Distributed PV Forecasting and Data Marketplace in an Era of Data Privacy Concerns

Ricardo Bessa, rbessa@inesctec.pt

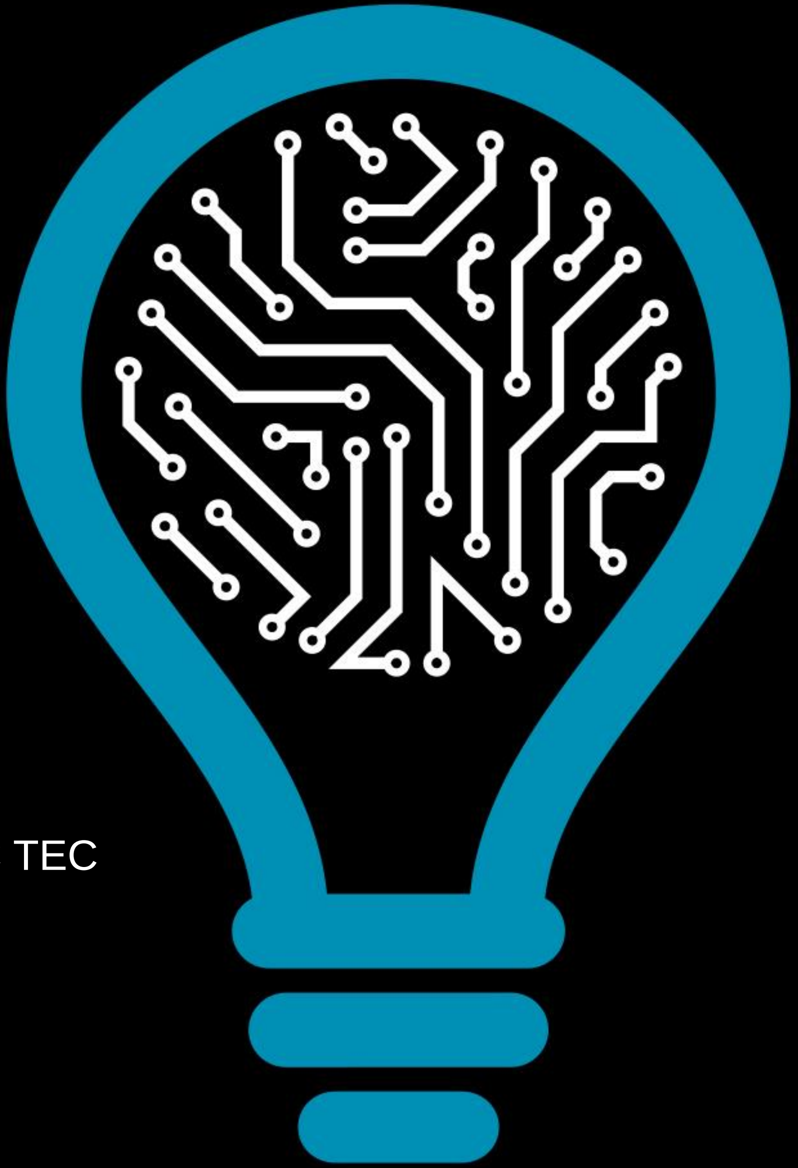
ESIG 2018 Forecasting Workshop - St. Paul, MN

20 June 2018

with Carla Gonçalves, Ricardo Andrade, Laura Cavalcante, INESC TEC

INESCTEC

INSTITUTE FOR SYSTEMS
AND COMPUTER ENGINEERING,
TECHNOLOGY AND SCIENCE



Key Messages

- **Information from a spatial grid of NWP improve forecasting skill**
- **PV sites can collaborate to improve forecasting skill and keep data private**
- **Grey box models can aggregate behind-the-meter information from flexible energy resources**

General Data Protection Regulation (GDPR) Hype

Opinion

Europe's Data Protection Law Is a Big, Confusing Mess

By Alison Cool
Ms. Cool is a professor of anthropology and information science at the University of Colorado, Boulder.

May 15, 2018

GDPR: Balancing Privacy And Innovation To Create Opportunities In Banking

Alan McIntyre, CONTRIBUTOR
FULL BIO
Opinions expressed by Forbes Contributors are their own.

The Startup · Follow · Sign in

HOME · EDITOR'S CHOICE · SUBMIT · GET SMARTER AT BUILDING STARTUPS

Data protection: Obsession or human right—will new policies kill AI innovation in Europe? 🤔

GDPR: The foundation for innovation

By Felix Marx 14 days ago Internet

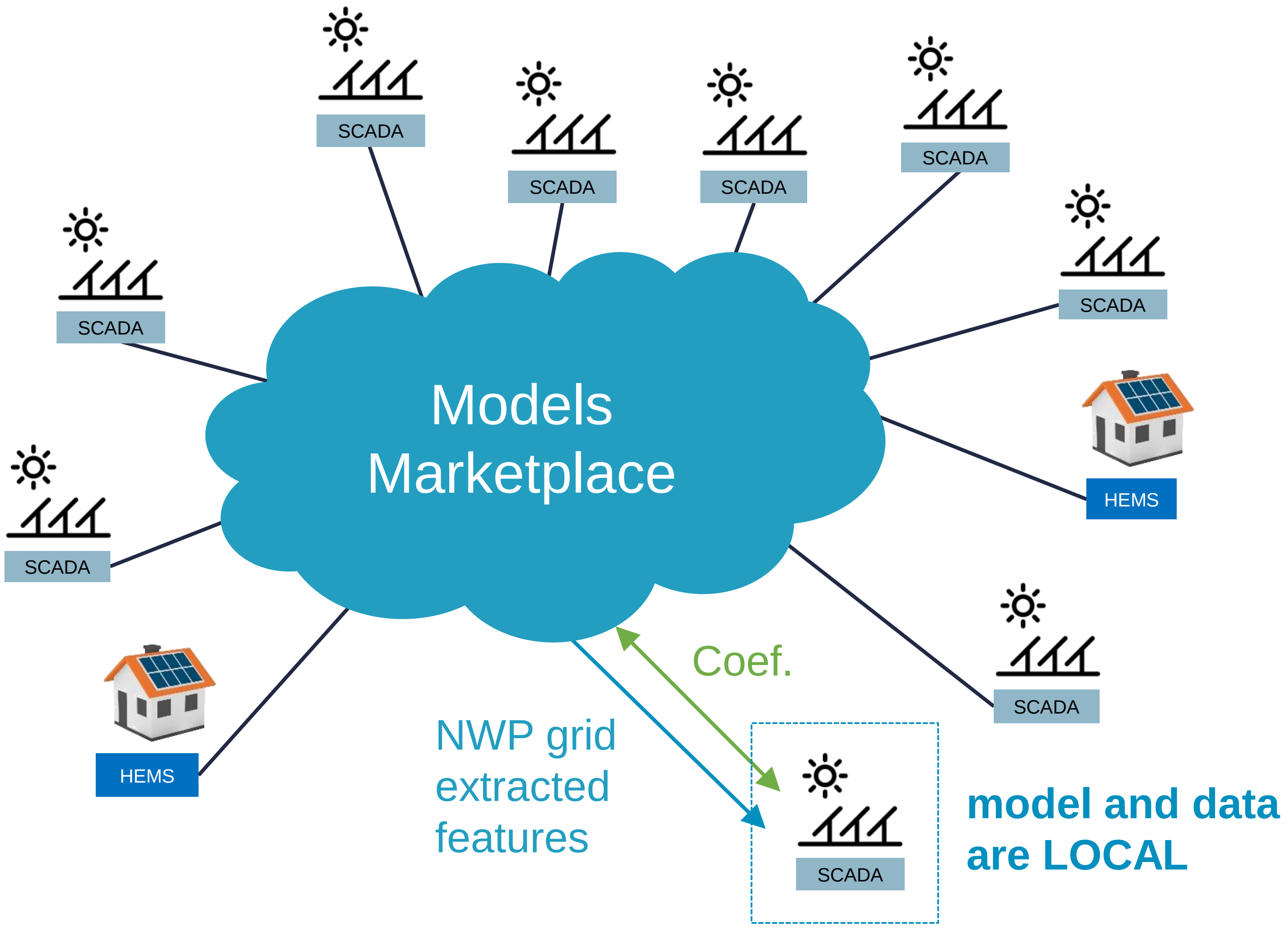
What benefits can GDPR bring for your business?

Re-think innovation towards distributed learning, federated learning and models marketplace

~~Data~~ Models Marketplace for PV Forecasting

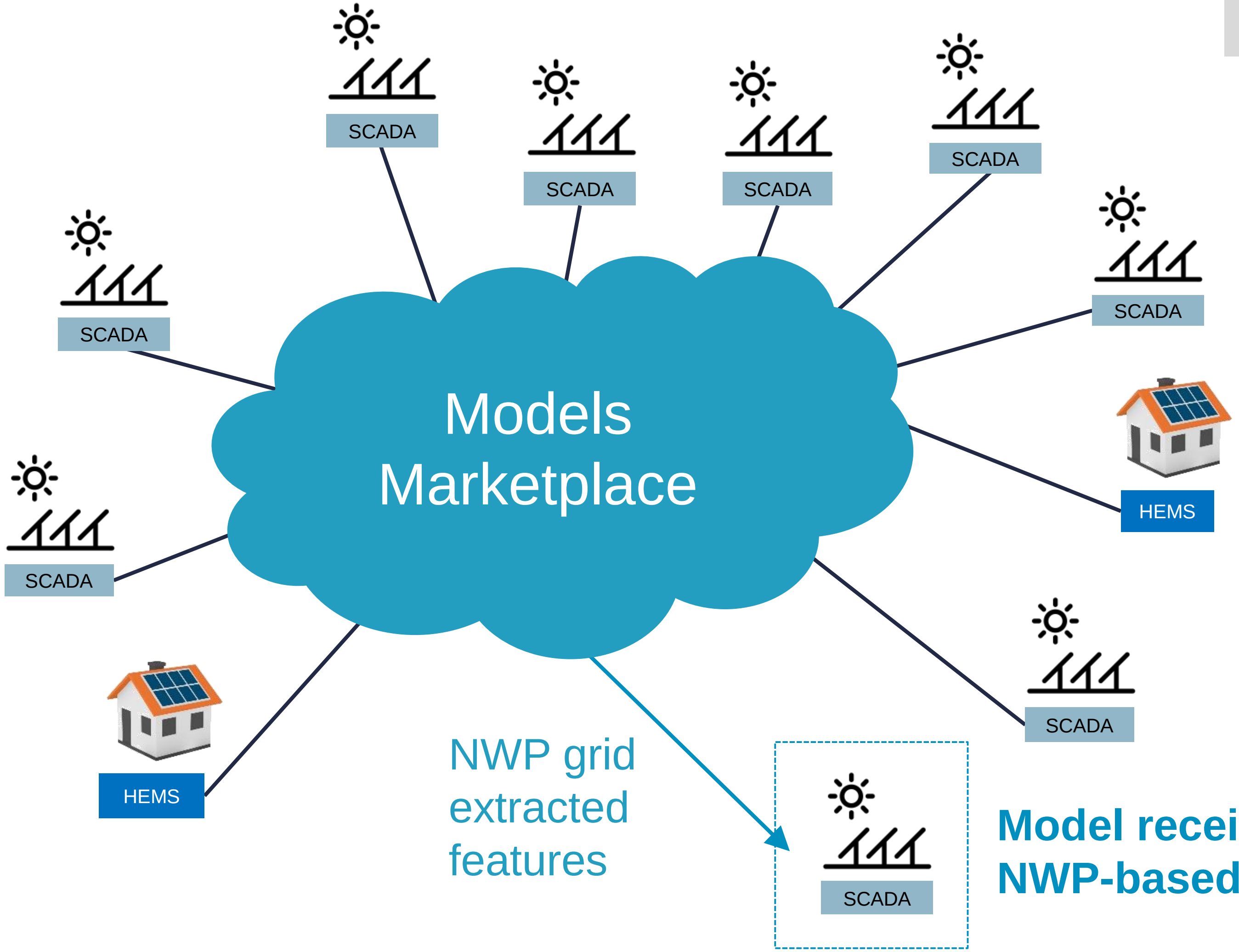
Main benefits

- Business case for collaborative (distributed) forecasting with data privacy
- Explore spatial-temporal measurements and NWP grid data to improve forecasting accuracy



Sell NWP Grid Data Features

day(s)-ahead horizon



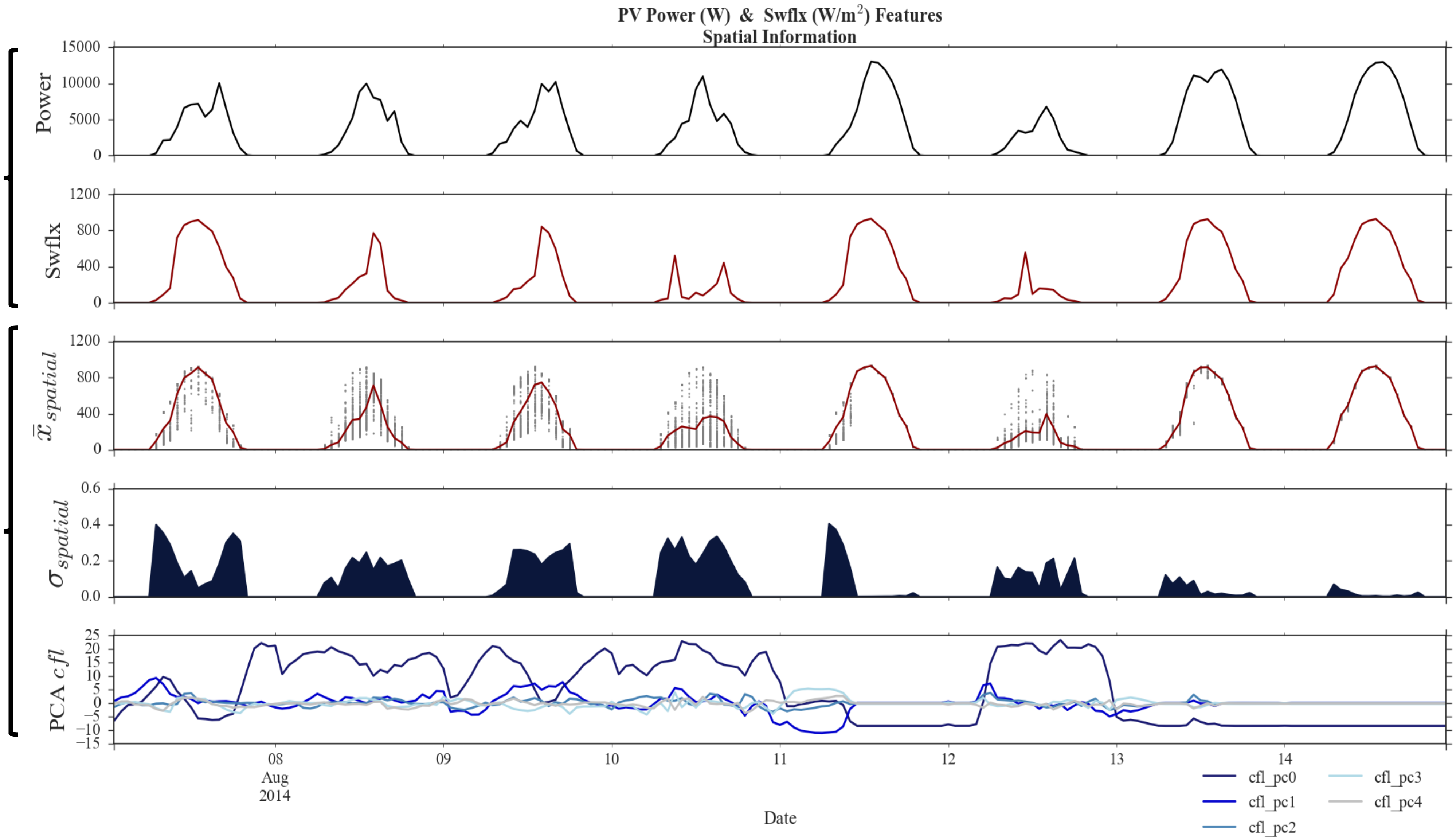
Model receives processed NWP-based features

J.R. Andrade, R.J. Bessa, "Improving renewable energy forecasting with a grid of numerical weather predictions", IEEE Transactions on Sustainable Energy, vol. 8, no. 4, pp. 1571-1580, Oct. 2017.

NWP Spatial Grid Features → PV Client

Local Information

Spatial Information



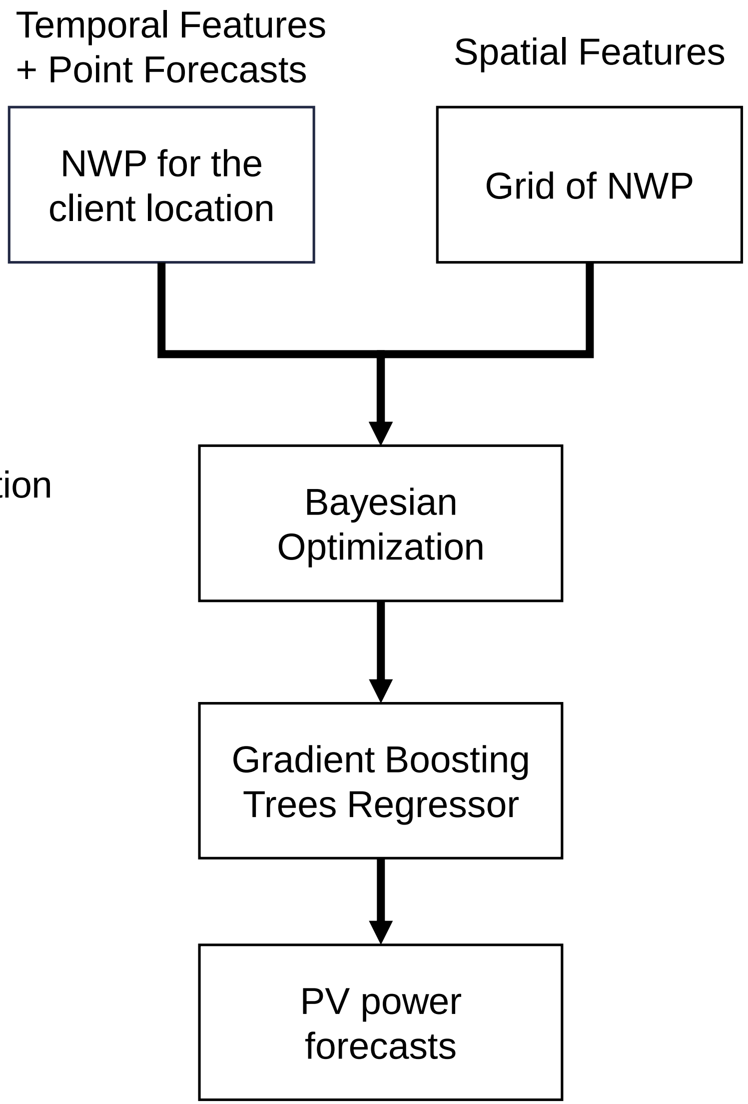
PV Client Statistical Model

- Weather forecasts

Constrained Global Optimization Solver:
- Parameter tuning
- Feature creation

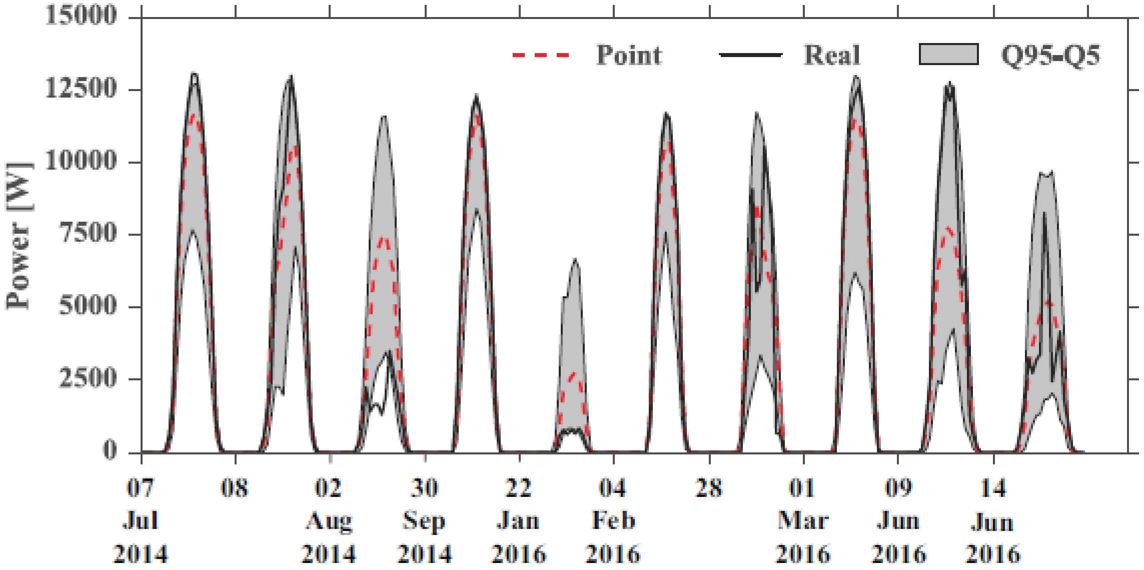
- Forecasting tool

- Point Forecasts
- Probabilistic Forecasts

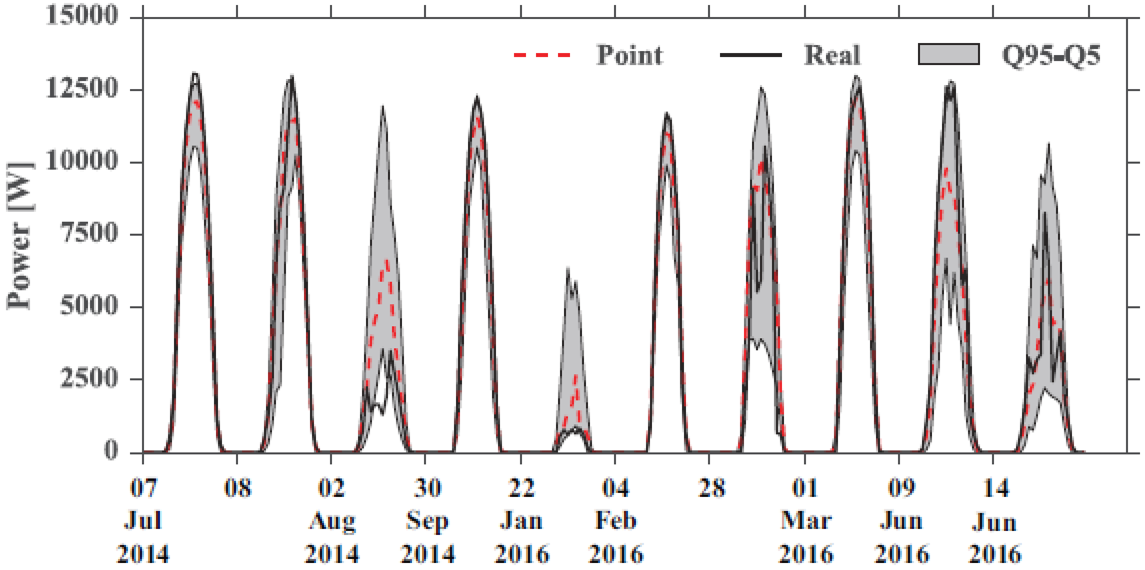


Illustrative Forecasts for a Building in Porto

Base model



Model: spatial & temporal data



Probabilistic forecasts:

- Uncertainty better modeled around the observed values
- “Abnormal” uncertainty verified in clear-sky days is removed

Point forecasts:

- Some of the over/underestimation situations are solved
- Improvements on the peak power forecasting in some clear-sky days

INESCTEC

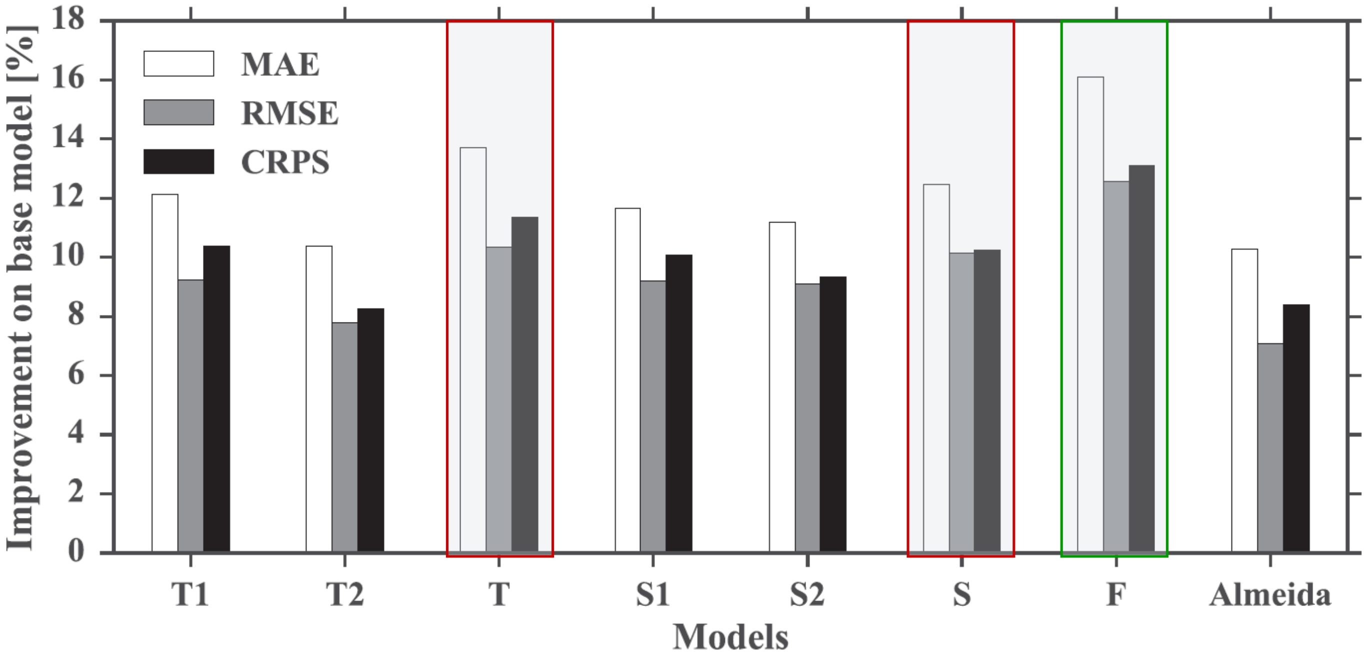
Forecasting Accuracy for a Building in Porto

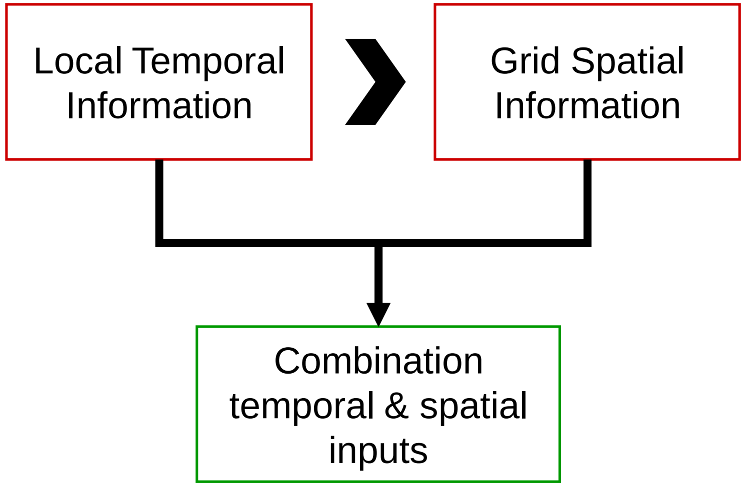
Base Model Inputs	
Chronological	Month Hour
NWP forecasts for the location of interest (INESC-TEC)	Surface downwelling shortwave flux [W/m^2] Temperature at 2m [K] Cloud cover at low levels [0, 1] Cloud cover at medium levels [0, 1] Cloud cover at high levels [0, 1] Cloud cover at low and medium levels [0, 1]

FEATURES CONSIDERED IN EVERY FORECASTING MODEL

Domain	ID	Features
Temporal	T1	Lags and leads
	T2	σ^2_{time} and different NWP runs
	T	Combination of models T1 and T2 inputs
Spatial	S1	$\sigma_{spatial}$ and $\bar{x}_{spatial}$
	S2	Principal components
	S	Combination of models S1 and S2 inputs
Temporal & Spatial	F	Combination of both domain features

24 hours ahead

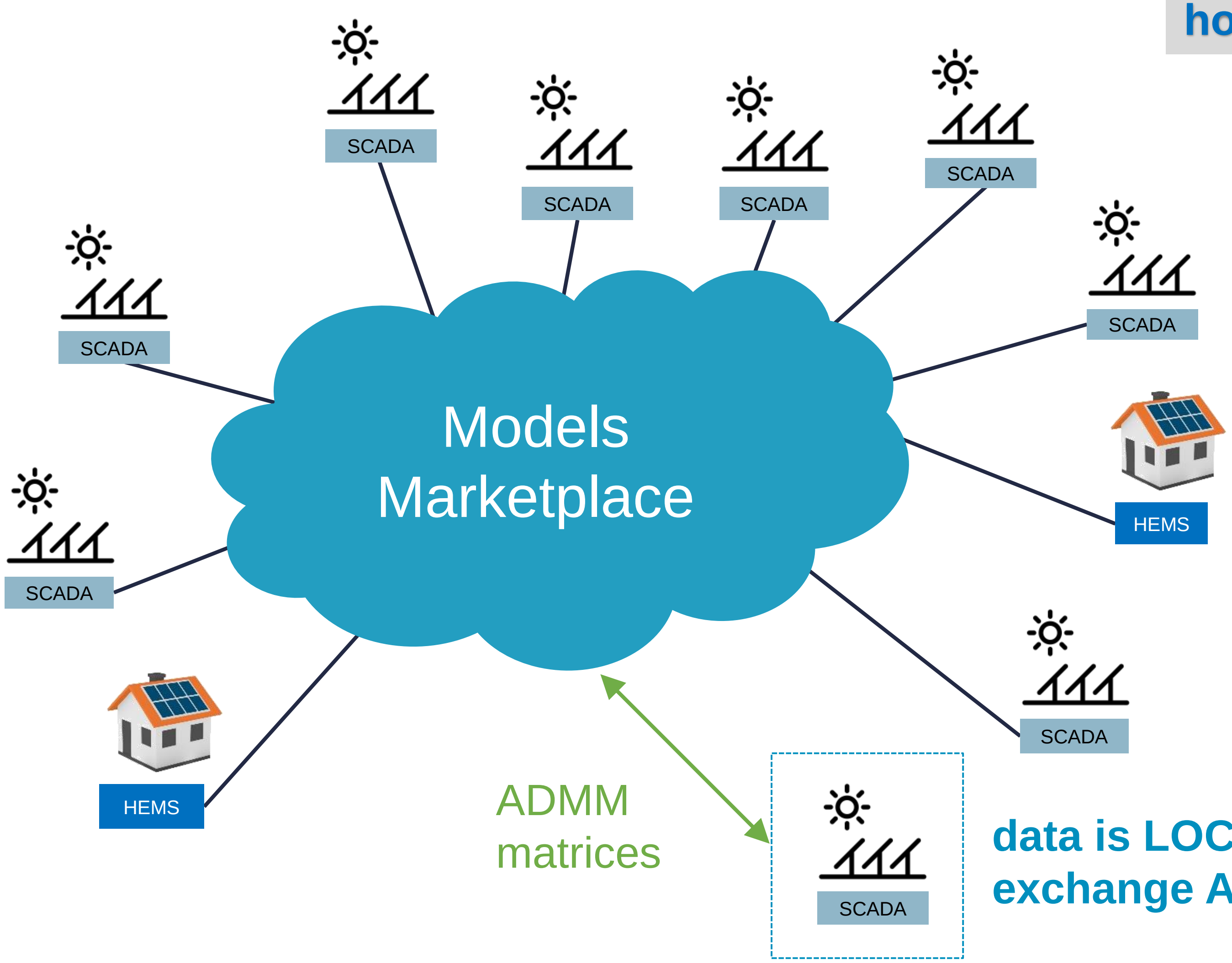




(best overall model)

Exchange Models Constructed with Distributed PV Measurements

hour(s)-ahead horizon



data is LOCAL and peers exchange ADMM matrices

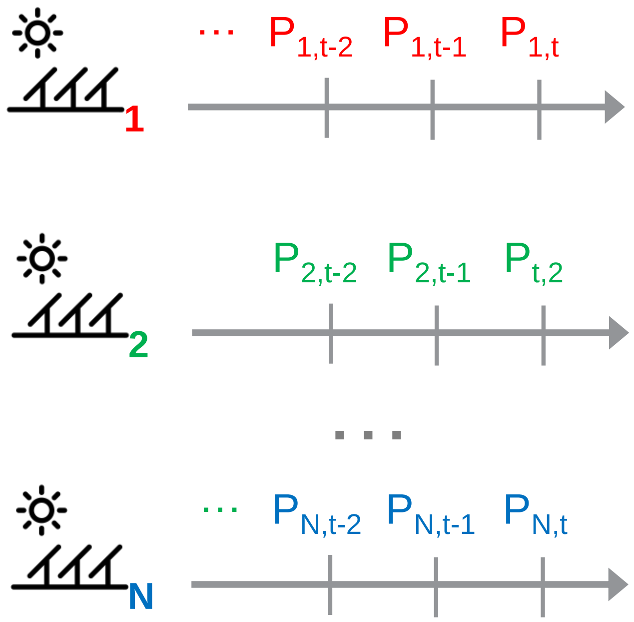
L. Cavalcante, R. J. Bessa, M. Reis, J. Dowell, "LASSO vector autoregression structures for very short-term wind power forecasting," Wind Energy, vol. 20, no. 4, pp. 657-675, April 2017.

VAR Model for Geographically Distributed PV Data

Example: matrix format for 2 PV sites

$$[P_{1,t} \quad P_{2,t}] = [c_1 \quad c_2] + \begin{bmatrix} B^1_{1,1} & B^1_{1,2} & B^2_{1,1} & B^2_{1,2} \\ B^1_{2,1} & B^1_{2,2} & B^2_{2,1} & B^2_{2,2} \end{bmatrix} \cdot \begin{bmatrix} P_{1,t-1} \\ P_{2,t-1} \\ P_{1,t-2} \\ P_{2,t-2} \end{bmatrix} + [E_{1,t} \quad E_{2,t}]$$

Z (lagged observations)



Vector Autoregressive Model (VAR)

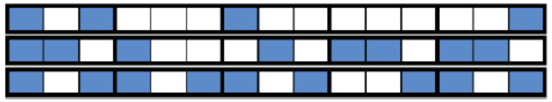
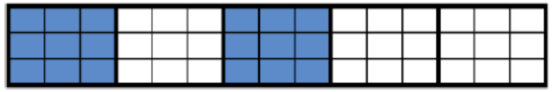
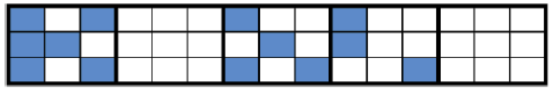
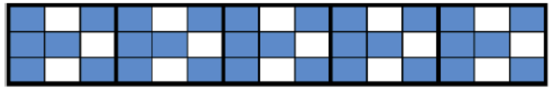
$$\boldsymbol{Y} = \boldsymbol{c} + \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{E}$$

Important Characteristics for the VAR Model

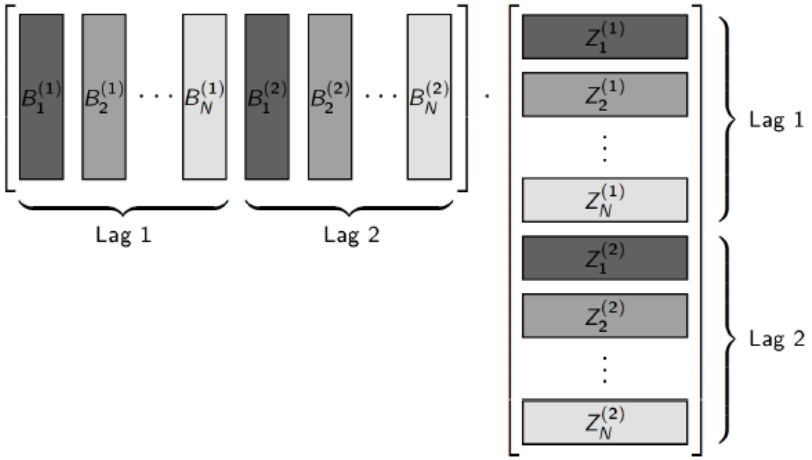
Sparse Coefficients Matrix (LASSO)

$$\frac{1}{2}\|Y - BZ\|_2^2 + \lambda\|B\|_1$$

Improves accuracy

LASSO-VAR Structures	Illustration
sLV	
rLV	
lLV	
lsLV	
ooLV	
cLV	

Distributed Learning



ADMM - alternating direction method of multipliers

Break up large datasets into blocks and carry out the VAR fitting over each block

Does not guarantee data privacy (→next slides)

Centralized LASSO-VAR Model

$$\boldsymbol{Y} = \boldsymbol{BZ} + \boldsymbol{E}$$

No private data shared between PV agents

VAR(2)

3 PV

1 … $P_{1,t-2}$ $P_{1,t-1}$ $P_{1,t}$

$\widehat{\boldsymbol{B}}_1^{(1)}, \widehat{\boldsymbol{B}}_1^{(2)}$

Neutral agent has direct access to private data

Neutral Agent

$$\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \boldsymbol{Y}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_1^{(1)} & \boldsymbol{B}_2^{(1)} & \boldsymbol{B}_3^{(1)} & \boldsymbol{B}_1^{(2)} & \boldsymbol{B}_2^{(2)} & \boldsymbol{B}_3^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_1^{(1)} \\ \boldsymbol{Z}_2^{(1)} \\ \boldsymbol{Z}_3^{(1)} \\ \boldsymbol{Z}_1^{(2)} \\ \boldsymbol{Z}_2^{(2)} \\ \boldsymbol{Z}_3^{(2)} \end{bmatrix}$$

ADMM LASSO-VAR

$$\hat{B} = \arg\min_B \frac{1}{2}\|Y - BZ\|_2^2 + \lambda\|B\|_1$$

$\widehat{\boldsymbol{B}}_2^{(1)}, \widehat{\boldsymbol{B}}_2^{(2)}$

$\widehat{\boldsymbol{B}}_3^{(1)}, \widehat{\boldsymbol{B}}_3^{(2)}$

2 … $P_{2,t-2}$ $P_{2,t-1}$ $P_{t,2}$

3 … $P_{3,t-2}$ $P_{3,t-1}$ $P_{3,t}$

Centralized LASSO-VAR Model

$$Y = BZ + E$$

No private data shared between PV agents

VAR(2)

3 PV

Transform the data using matrix multiplication

Matrix Q invertible and such that $QQ^T = I$

Defined and shared between PV agents

… $P_{1,t-2}$ $P_{1,t-1}$ $P_{1,t}$

1

$\widehat{B}_1^{(1)}, \widehat{B}_1^{(2)}$

$$Y_1 = (P_{1,t} \quad \cdots \quad P_{1,t+h})Q$$
$$Z_1^{(1)} = (P_{1,t-1} \quad \cdots \quad P_{1,t+h-1})Q$$
$$Z_1^{(2)} = (P_{1,t-2} \quad \cdots \quad P_{1,t+h-2})Q$$

Neutral Agent

…

$\widehat{B}_2^{(1)}, \widehat{B}_2^{(2)}$

$$Y_2 = (P_{2,t} \quad \cdots \quad P_{2,t+h})Q$$
$$Z_2^{(1)} = (P_{2,t-1} \quad \cdots \quad P_{2,t+h-1})Q$$
$$Z_2^{(2)} = (P_{2,t-2} \quad \cdots \quad P_{2,t+h-2})Q$$

$\widehat{B}_3^{(1)}, \widehat{B}_3^{(2)}$

$$Y_3 = (P_{3,t} \quad \cdots \quad P_{3,t+h})Q$$
$$Z_3^{(1)} = (P_{3,t-1} \quad \cdots \quad P_{3,t+h-1})Q$$
$$Z_3^{(2)} = (P_{3,t-2} \quad \cdots \quad P_{3,t+h-2})Q$$

2

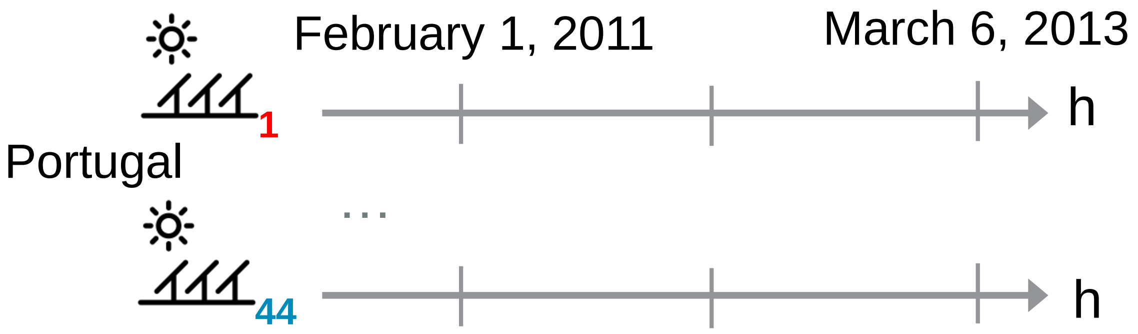
… $P_{2,t-2}$ $P_{2,t-1}$ $P_{t,2}$

3

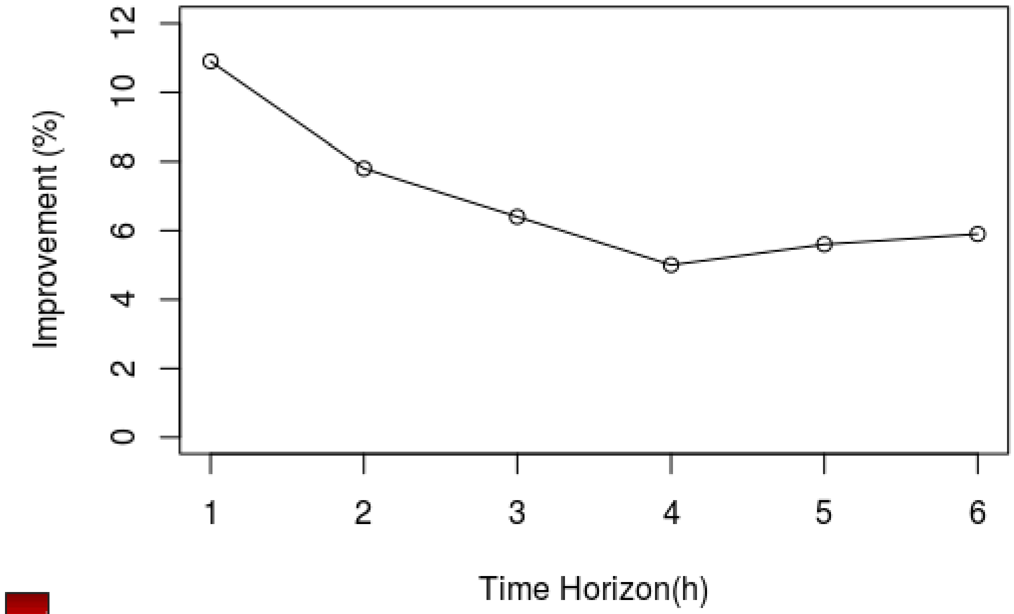
… $P_{3,t-2}$ $P_{3,t-1}$ $P_{3,t}$

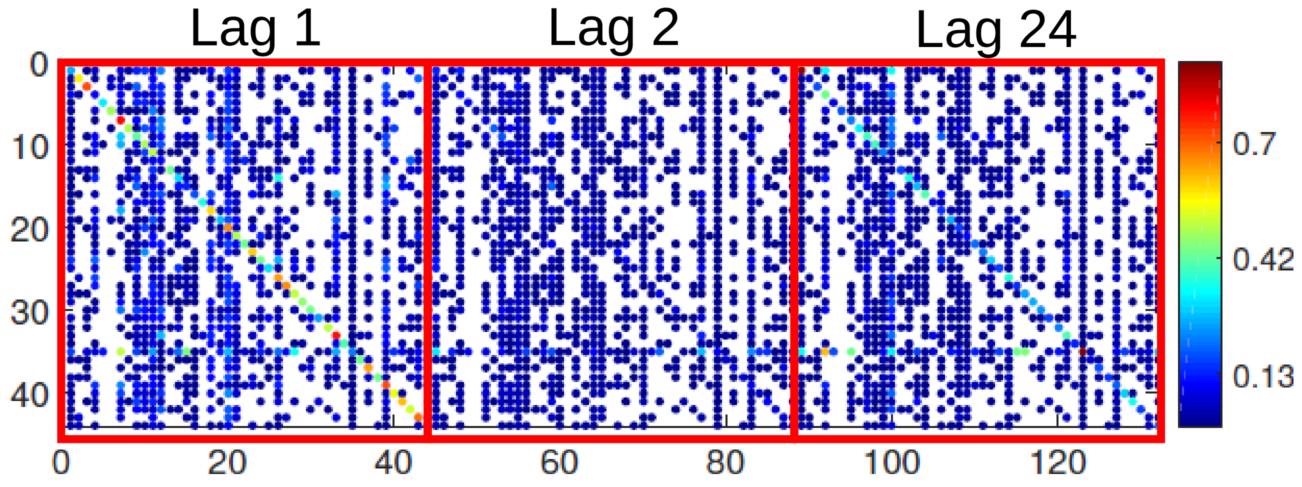
If $QQ^T = I$ then the ADMM VAR-LASSO solution for $Y = BZ$ and $YQ = BZQ$ remains the same

Results for a Smart Grid Pilot



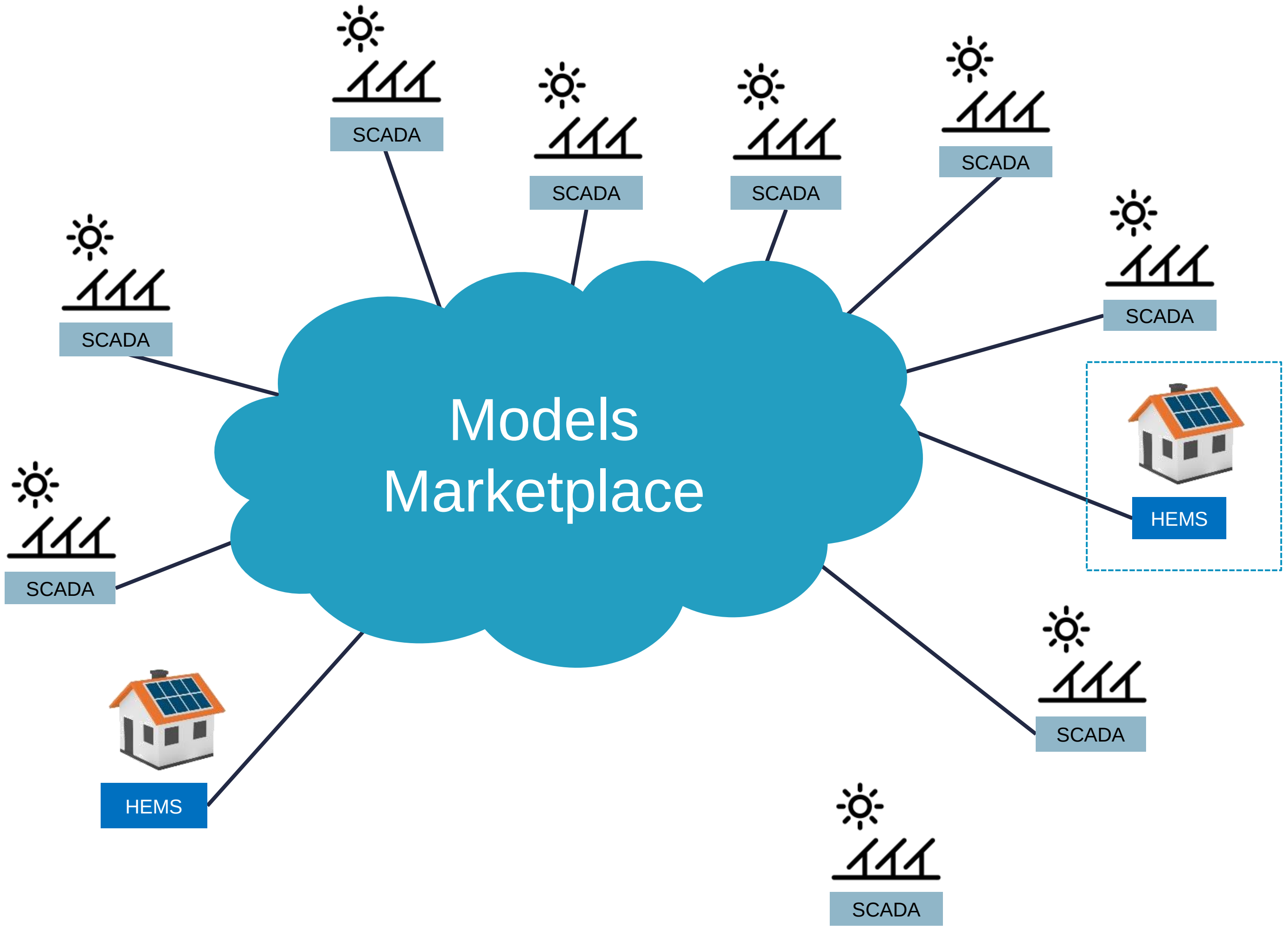
RMSE improvement (%) of the cLV structures using a sliding window approach over AR model.





Coefficients matrices of the cLV structure for first lead-time (lags 1,2 and 24 h)

Behind-the-Meter Flexibility

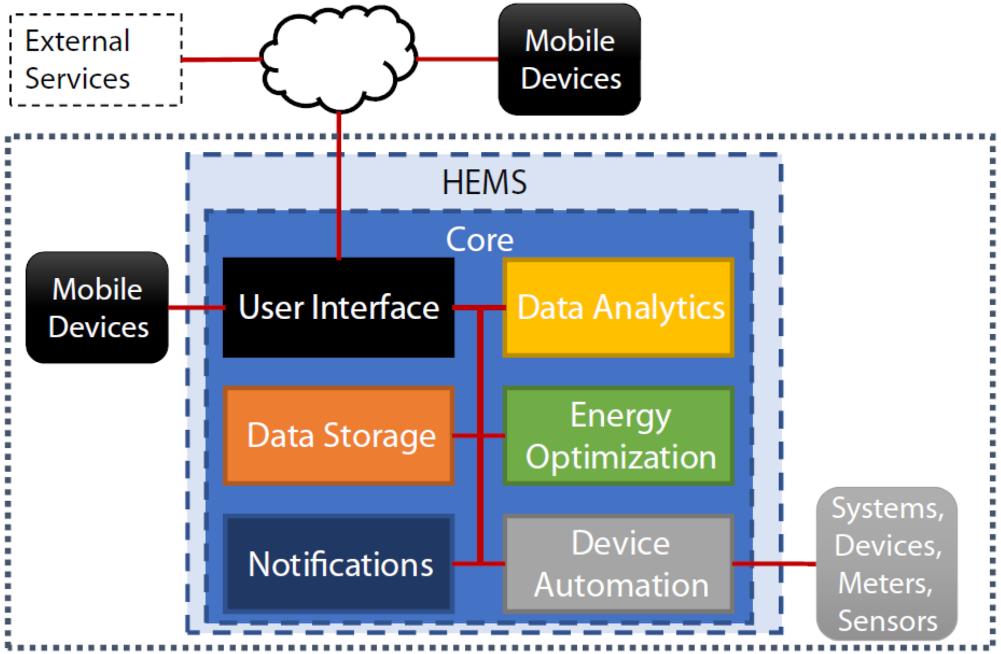


Communicate flexibility with machine learning OR virtual dynamic batteries

R.J. Bessa, D. Rua, C. Abreu, P. Machado, J.R. Andrade, R. Pinto, C. Gonçalves, and M. Reis, "Data economy for prosumers in a smart grid ecosystem," in Proc. of the e-Energy ’18: The Nineth International Conference on Future Energy Systems, June 12–15, 2018, Karlsruhe, Germany.

R.B. Pinto, R.J. Bessa, M.A. Matos, "Multi-period flexibility forecast for low voltage prosumers," Energy, vol. 141, pp. 2251-2263, Dec. 2017.

Energy Management & Flexibility Trajectories



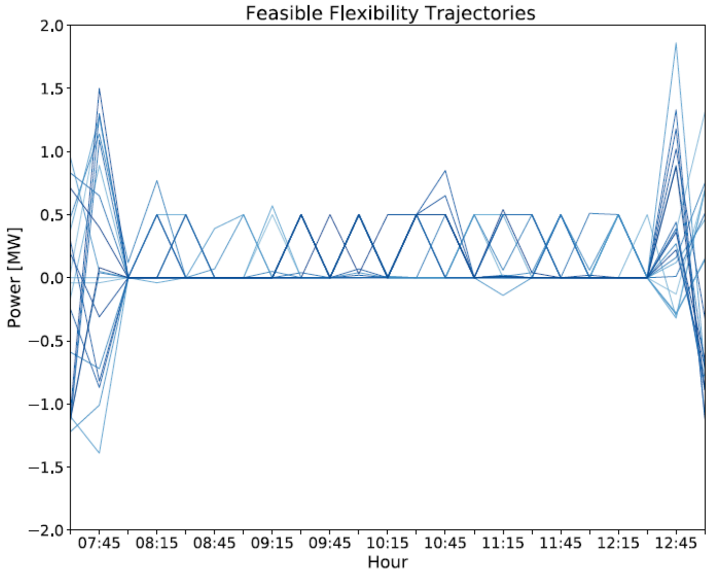
Home Energy Management System (HEMS)

PV uncertainty forecasting

Load uncertainty forecasting

Storage & electric water heaters technical characteristics

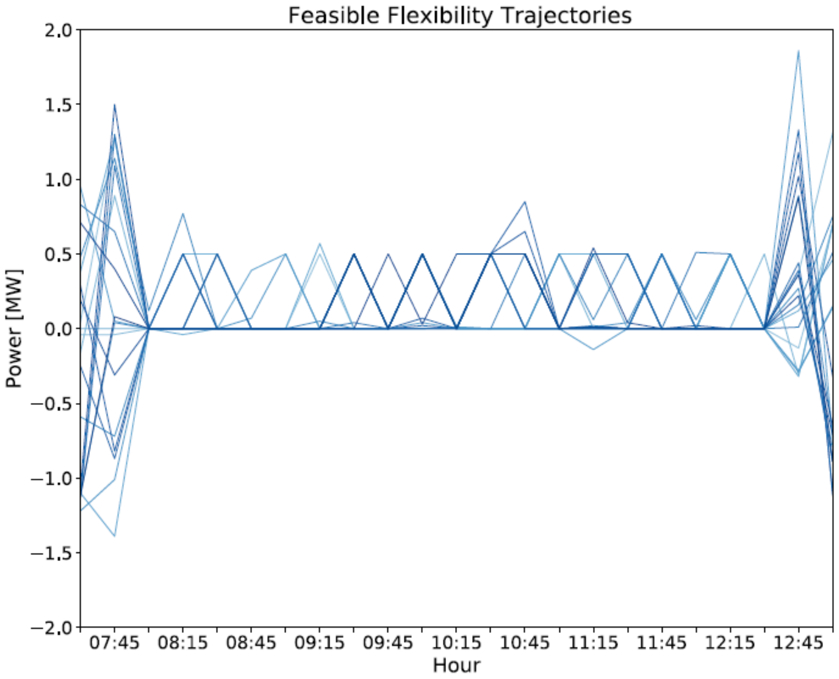
Feasible trajectory search algorithm EPSO



integrid bridging the gap

This project is funded by the European Union

Flexibility as a Data-Driven Model



two products

P1

support vector data description (SVDD)

$$R^2(x) = 1 - 2\sum_i \beta_i \cdot k(x_i, x) + \sum_{i,j} \beta_i \cdot \beta_j \cdot k(x_i, x_j)$$

P2

Dynamic "virtual" battery

- ► SOC^{max} vary. with time t
- ► P^{max} vary. with time t

Sell flexibility models instead of exchanging behind-the-meter data from prosumers

Concluding Remarks

Opportunities & Challenges

MODEL-BASED SERVICES

Trade models, instead of data or forecasting services

EMBEDDED SYSTEMS

“Light” distributed and online statistical learning algorithms

SCALABILITY

Peer-to-peer schemes with asynchronous communication

FEATURE EXTRACTION

can lead to significant forecasting skill improvement

STANDARDIZATION

Standards are needed…!

PRIVACY-PRESERVING ANALYTICS

Data-driven models compatible with GDPR and client concerns

- INESC TEC
- R DR. ROBERTO FRIAS
- 4200-465 PORTO
- PORTUGAL

T +351 222 094 000
F +351 222 094 050
info@inesctec.pt
www.inesctec.pt